

THE KEPLER DATA PROCESSING HANDBOOK: A FIELD GUIDE TO PROSPECTING FOR HABITABLE WORLDS

Jon M. Jenkins, NASA Ames Research Center and the *Kepler* Science Operations Center and Science Office Teams

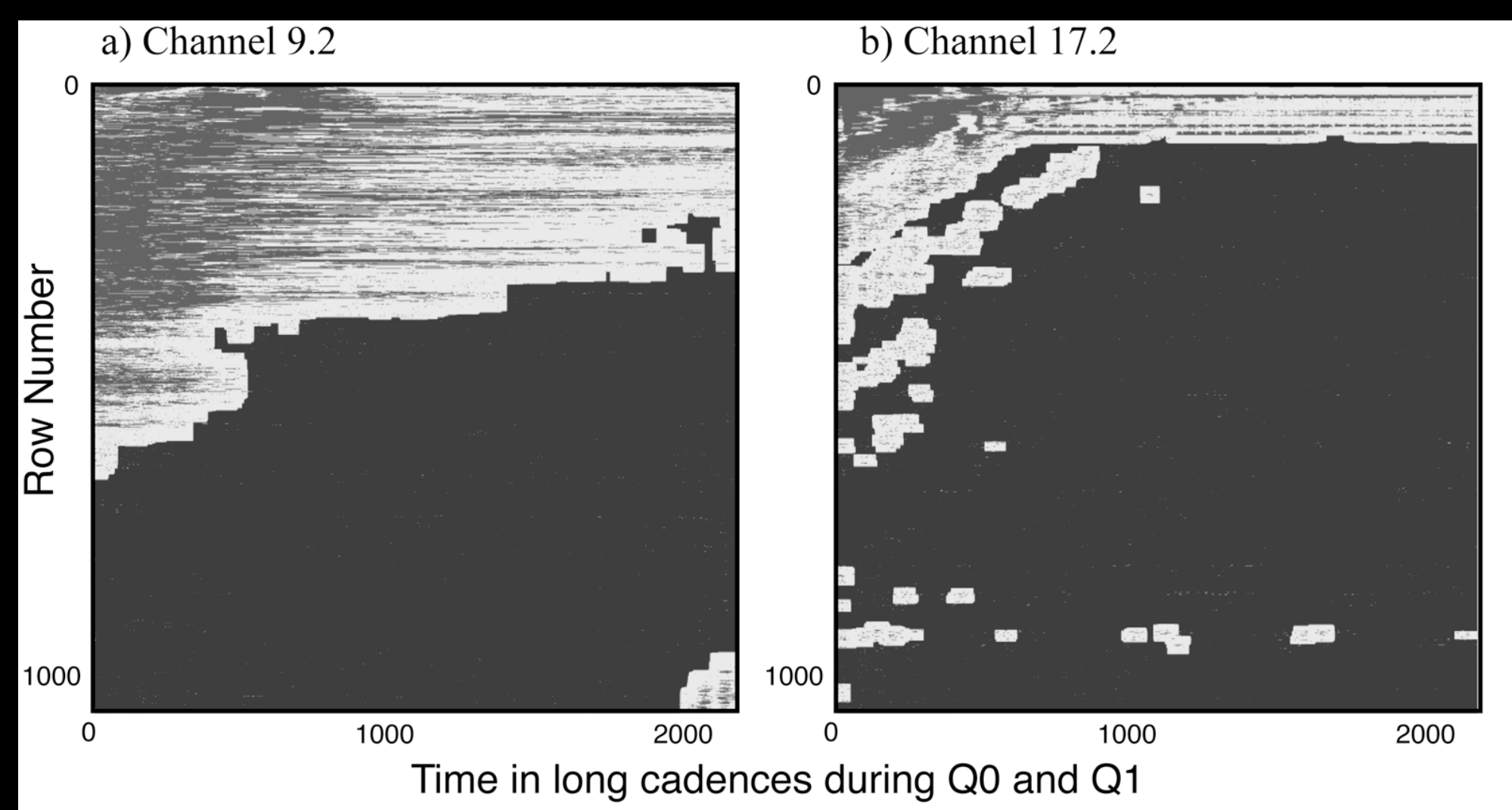
ABSTRACT

The Kepler telescope hurtled into orbit in March 2009, initiating NASA's first mission to discover Earth-size planets orbiting Sun-like stars. Kepler simultaneously collected data for ~165,000 target stars at a time over its four-year mission, identifying over 4700 planet candidates, over 2300 confirmed or validated planets, and over 2100 eclipsing binaries. While Kepler was designed to discover exoplanets, the long-term, ultra-high photometric precision measurements it achieved made it a premier observational facility for stellar astrophysics, especially in the field of asteroseismology, and for variable stars, such as RR Lyrae. The Kepler Science Operations Center (SOC) was developed at NASA Ames Research Center to process the data acquired by Kepler from pixel-level calibrations all the way to identifying transiting planet signatures and subjecting them to a suite of diagnostic tests to establish or break confidence in their planetary nature. Detecting small, rocky planets transiting Sun-like stars presents a variety of daunting challenges, including achieving an unprecedented photometric precision of ~20 ppm on 6.5-hour timescales, and supporting the science operations, management, processing, and repeated reprocessing of the accumulating data stream.

A newly revised and expanded version of the Kepler Data Processing Handbook (KDPH) has been released to support the legacy archival products. The KDPH details the theory, design and performance of the algorithms supporting each data processing step. This paper presents an overview of the KDPH and features illustrations of several key algorithms in the Kepler Science Data Processing Pipeline.

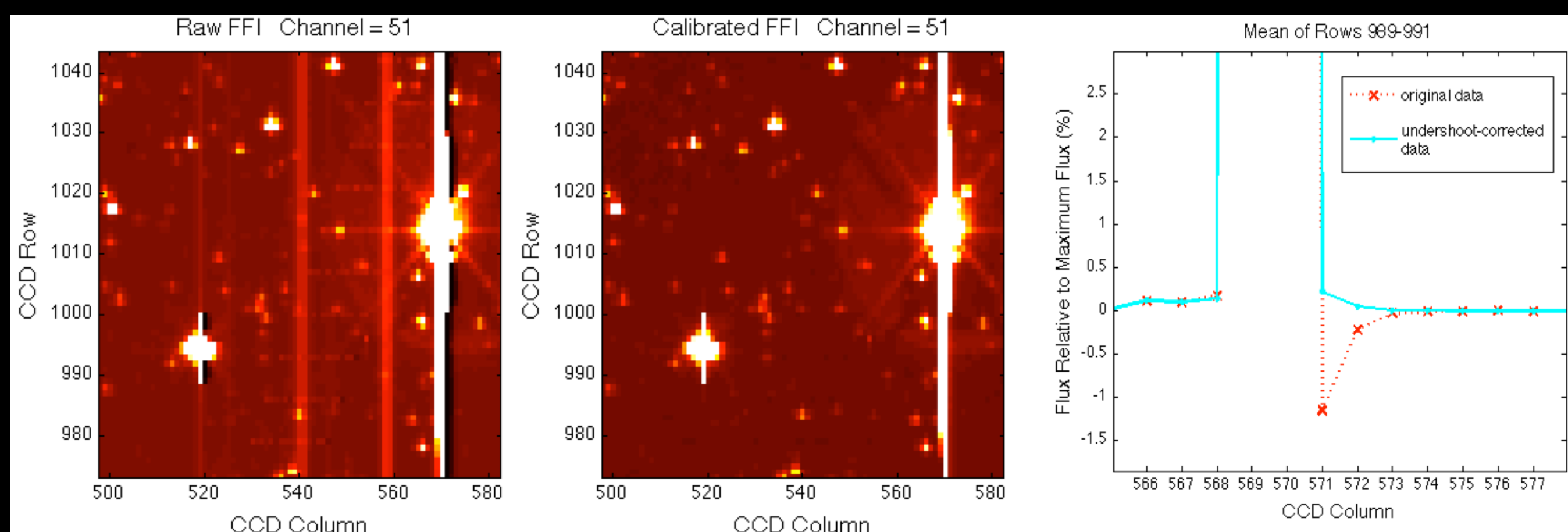
Kepler was selected as the 10th mission of the Discovery Program. Funding for this mission is provided by NASA, Science Mission Directorate.

DYNABLACK



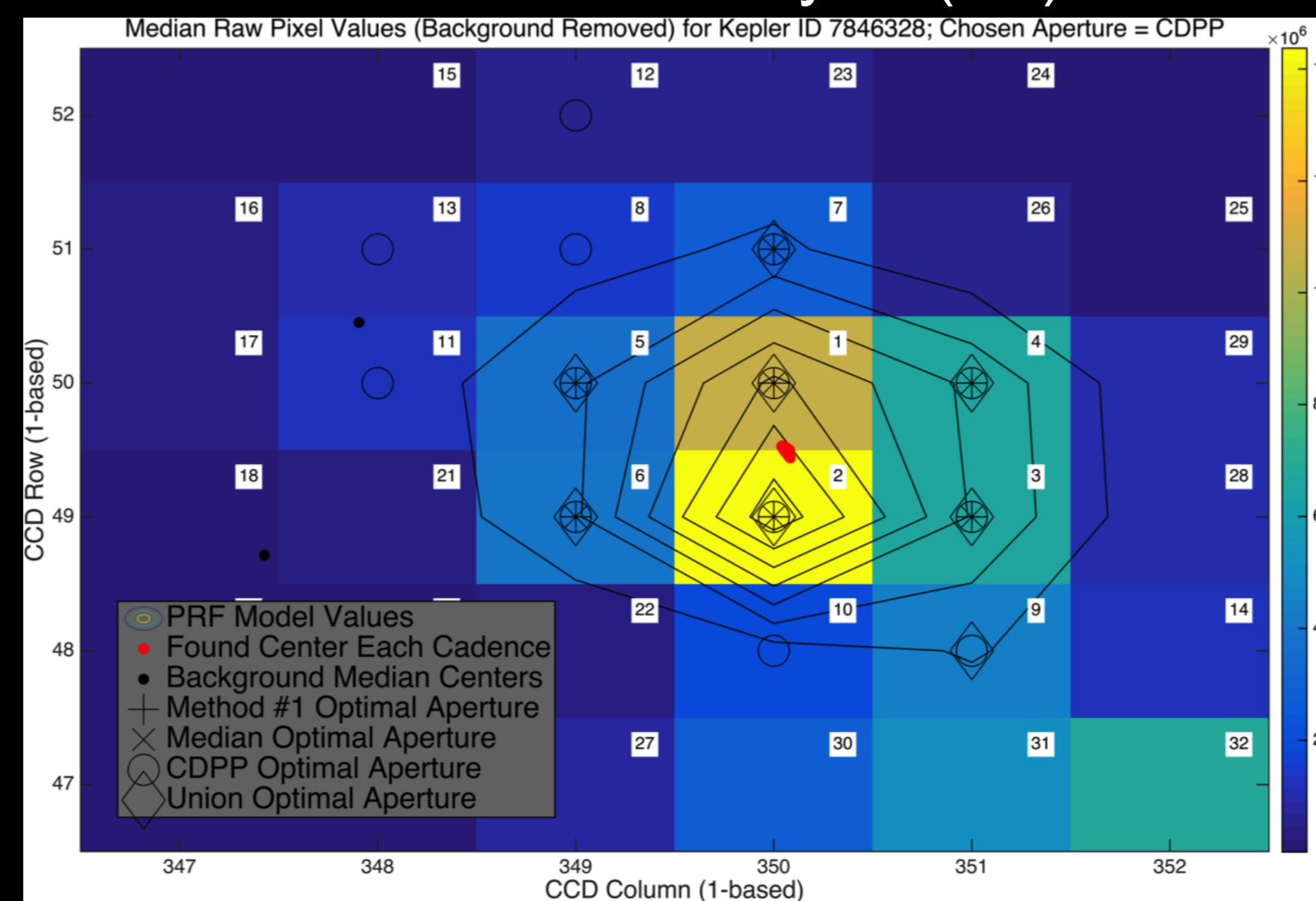
Example of rolling band flagging results from channels a) 9.2 and b) 17.2 for combined Q0-Q1. Varying shaded areas represent the degree beyond threshold level that the data indicate a rolling band, white regions are the padding around the offending regions and black areas are unflagged. The scales are image row vs. time as measured in long cadence periods. From Figure 7 of Kolodziejczak et al. (2010).

CALibration



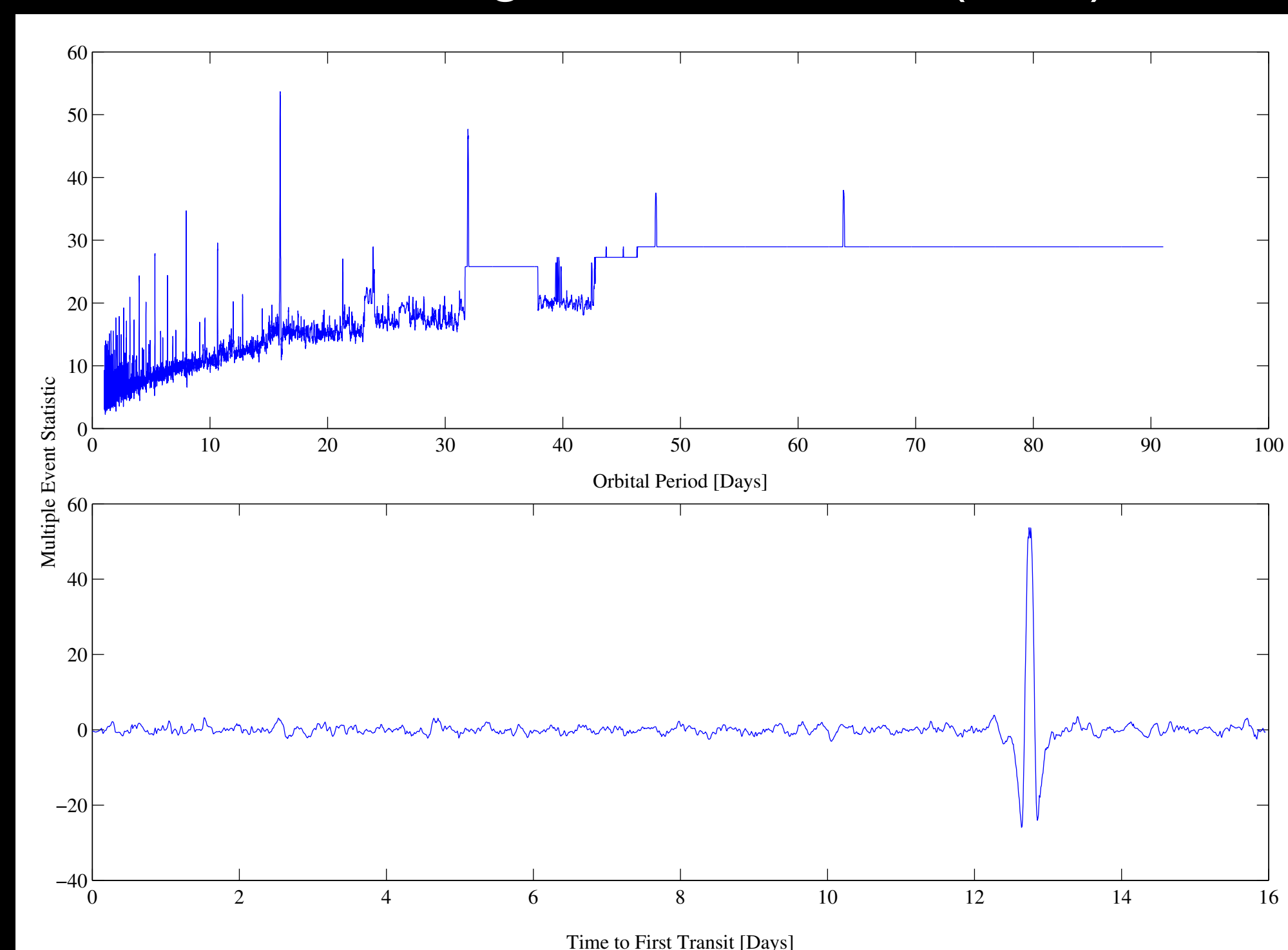
A close up, stretched image of two saturated target stars that show pixel undershoot signatures resulting from bright to dark pixel transitions in the direction of the serial readout (left panel) and the calibrated image (middle). The mean of 3 pixel rows is shown for one target (right) with the undershoot response (the negative spike) along with the corrected pixel values. From Figure 7 of Quintana et al. (2010).

Photometric Analysis (PA)



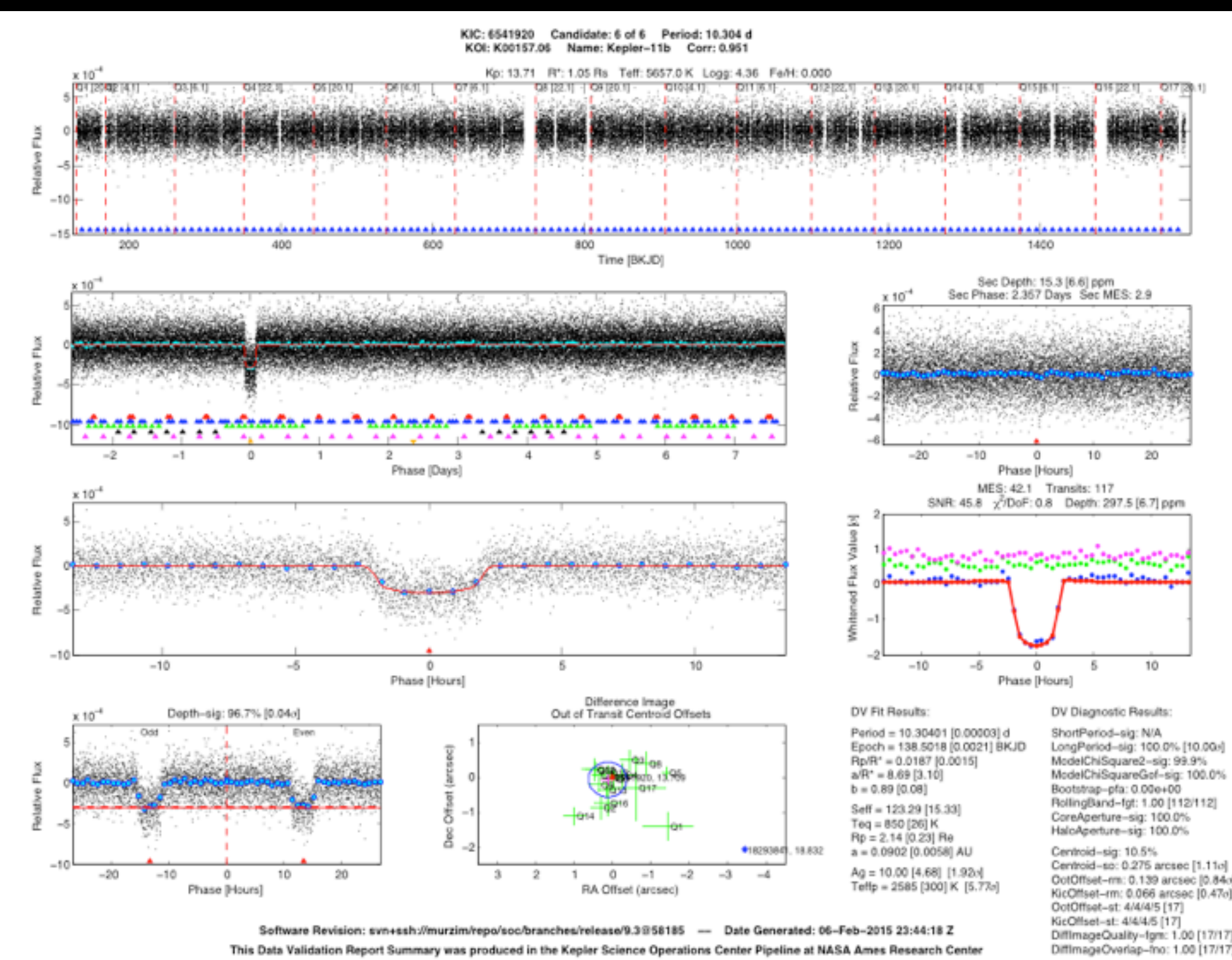
The mask scene, fitted PRF image and four found apertures for Kepler ID 7846328. The color bar is in units of electrons per second. From Figure 3 in Smith et al. (2016).

Transiting Planet Search (TPS)



Multiple event statistics (MES) determined by folding the single event statistics distribution. Top: Maximum multiple event statistic as a function of fold interval (orbital period), showing a peak at 15.97 days, corresponding to the orbital period of the transiting object in the data of Figure 9.7. Bottom: MES as a function of lag time for a 15.97 day period, showing a peak at 12.74 days, corresponding to the midtime of the first transit shown in Figure 9.7. From Figure 8 of Jenkins et al. (2010b)

Data Validation (DV)



A one-page summary of transiting planet model fit and diagnostic test results is produced for each planet candidate identified in a TPS/DV pipeline run. The summaries are delivered to MAST along with the full reports where they are accessible by the community at large. The summary above was produced for Kepler-11b with the final version of the SOC codebase, SOC-9.3.

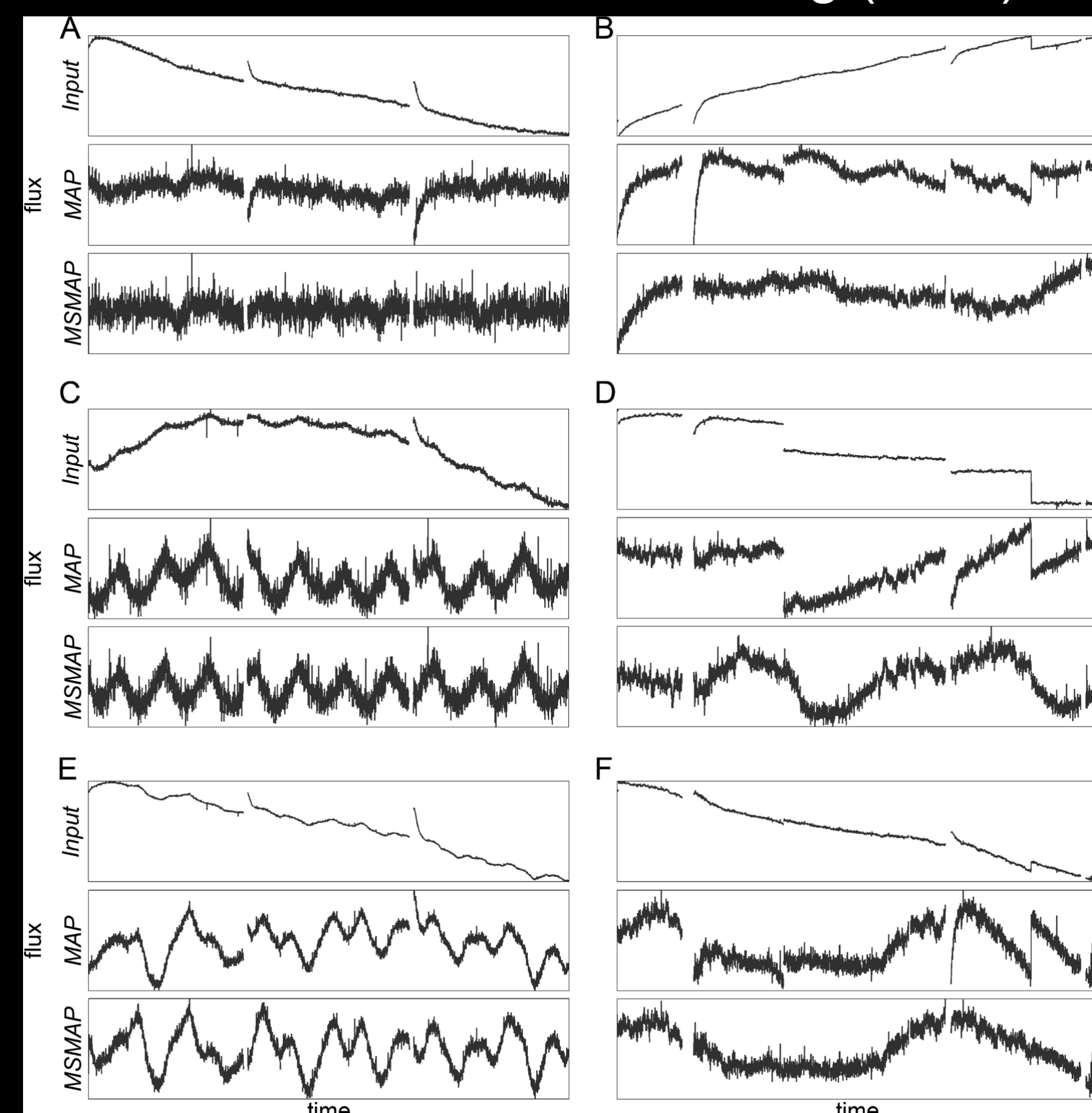
KEPLER SCIENCE DATA PROCESSING PIPELINE



KEPLER&K2
SciConIV

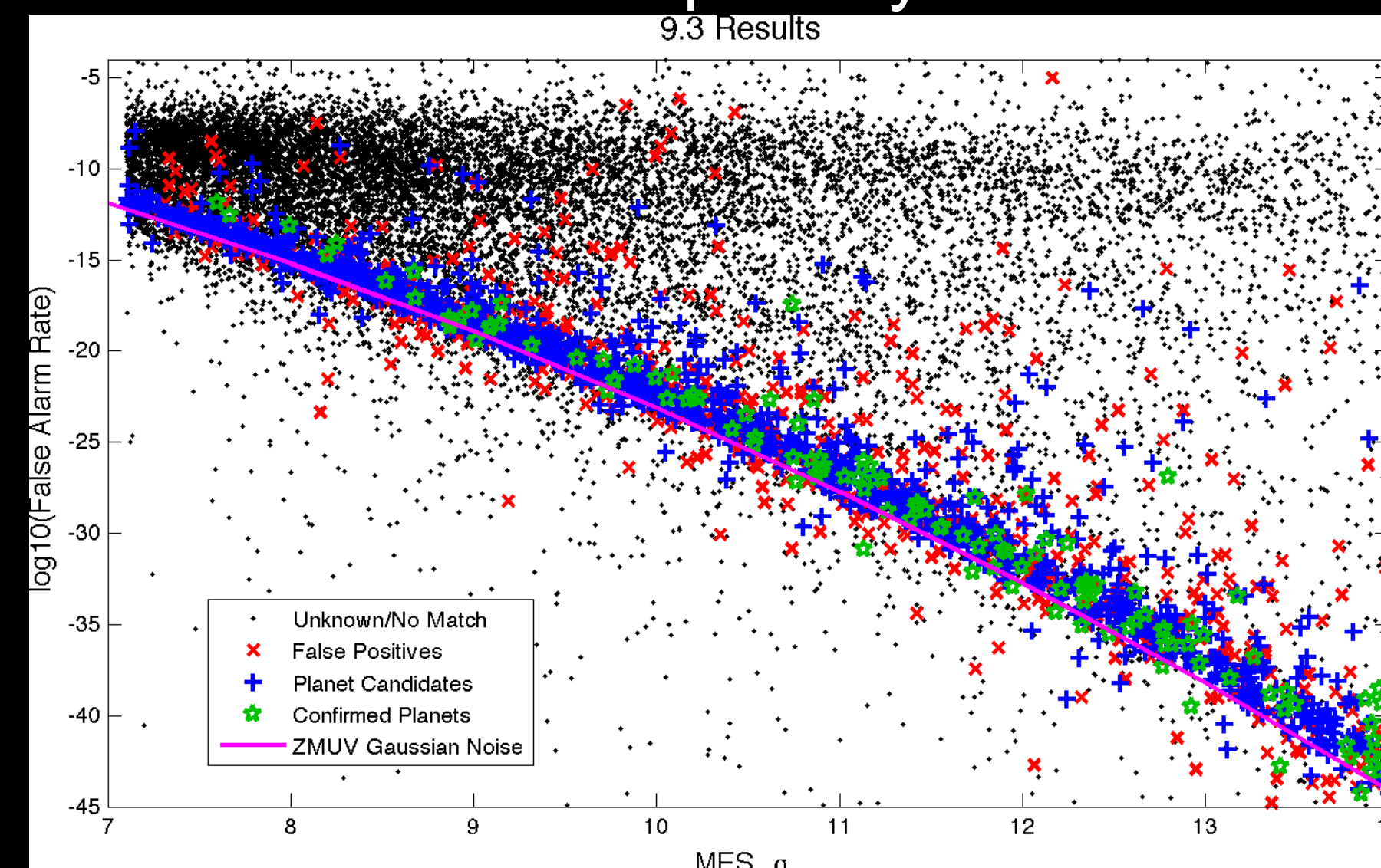
June 19—June 23, 2017

Presearch Data Conditioning (PDC)



Examples of msMAP correction performance improvements. The top row in each panel shows the PDC input time series, the middle row the PDCMAP corrected time series, and the bottom row the msMAP corrected time series. Vertical scales are rescaled between panels to show detail. Panels A, C, E show examples with regular data quality (Q10 data), and panels B, D, F show cases with discontinuities due to attitude tweaks (Q2 data). The example in panel A and B are the same ones as in Figure 8.39. From Figure 1 of Stumpe et al. (2014). From Figure 14 of Stumpe et al. (2014).

Bootstrap Analysis



False alarm rate as a function of the MES for each of the SOC 9.3 Q1-Q17 DR25 TCEs, colored by KOI disposition.